Instance search, Copy detection, Semantic indexing @ TRECVID

NIST
National Institute of Standards and Technology

What is TRECVID?

NIST workshop series (2001 – present) → http://trecvid.nist.gov

Foundation for large-scale laboratory testing on video

Forum for the

- exchange of research ideas and for
- the discussion of research methodology – what works, what doesn't , and why.

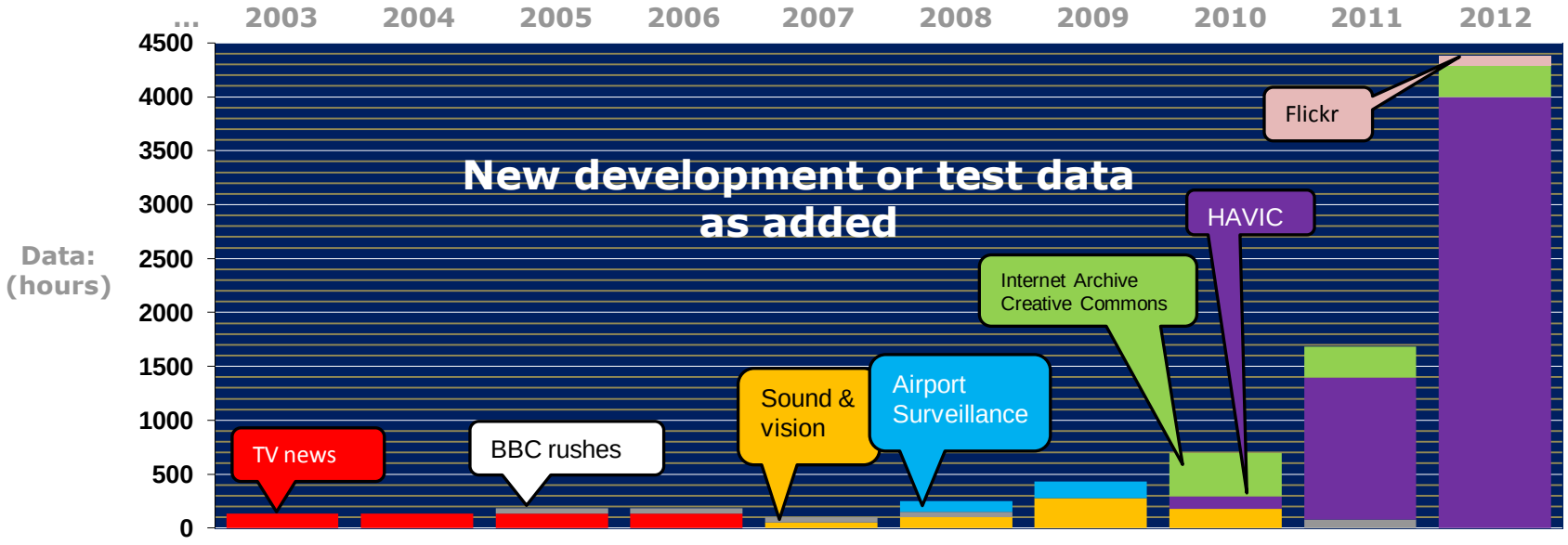
Focus: content-based analysis, retrieval,detection, summarization, etc.

Aims for realistic system tasks and test collections

- unfiltered data
- focus on relatively high-level functionality (e.g. interactive search)
- measurement against human abilities

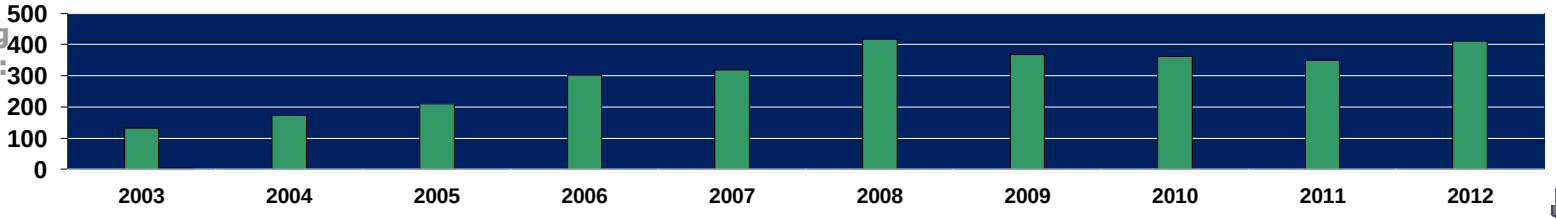
Provides data, tasks, and uniform, appropriate scoring procedures

TRECVID's Evolution



Tasks:

```
Ad hoc search
Features/semantic indexing
Stories                  Camera motion
BBC rushes -  -  -  -                        Summaries
Copy detection -  -  -  -  -  -  -  -  -  -  -  -
Surveillance events (SED)  -  -  -  -  -  -  -  -
Known-item search (KIS) -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -
Instance search pilot (INS) -  -  -  -  -  -  -  -  -  -  -  -  -  -  -
Multimedia event detection (MED) -  -  -  -  -  -  -  -  -  -  -  -  -
Multimedia event recounting (MER) -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -
```

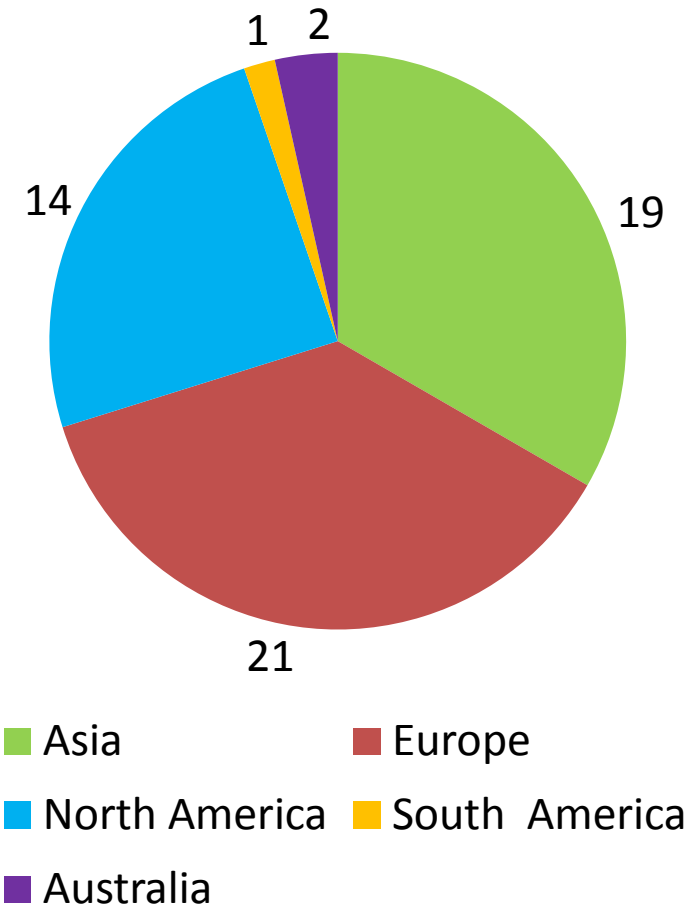


TRECVID 2012 Tasks and Data

Internet Archive – Creative Commons (IACC) *291 hours test data,video, title, keywords, description* *687 hours dev video, metadata*	*Flickr* *91 hours, test video clips divided into 76751 files*	*Airport surveillance* *50 hours, test video* *100 hours, dev video*	*HAVIC - Internet multimedia* *3722 hours, test video* *1429 hours, dev video*
Known-item search (361 text-only ad hoc test queries)	**Instance search pilot** (21 ad hoc image test queries)	**Surveillance event detection (interactive)** (7 known test events, choose 3)	**Multimedia event detection** (20 pre-specified & 5 ad hoc complex test events)
Semantic indexing (346 known test concepts, auto-assigned, submitted; 46 evaluated; 10 concept-pairs evaluated)			**Multimedia event recounting pilot** (5 events)

TV2012 Finishers

Groups Finished	Task code	Task name
13	SED	Surveillance event detection
25	SIN	Semantic indexing
9	KIS	Known-item search
24	INS	Instance search pilot
17	MED	Multimedia event detection
10	MER	Multimedia event recounting



Semantic indexing

Example use case: *automatically filter a very large collection of video to find just the clips which contain examples of a given* ***concept class.***

2012 System task:

- Given 356 concept names and a test collection of video 145,634 clips (200 hours),
- For each concept, return a list of up to 2000 clips ranked by the likelihood they contain the target concept

Applications:

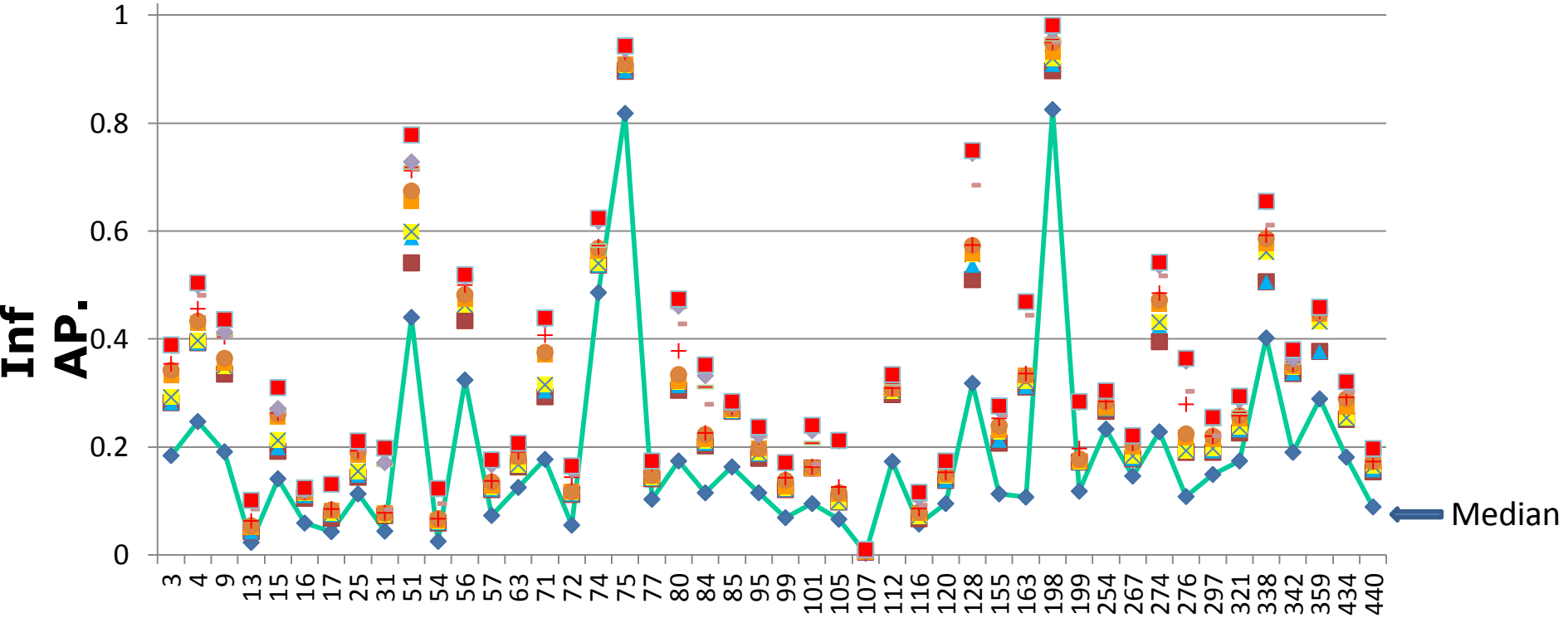
- Search: used to execute more complicated video queries by breaking them into pieces
- Law enforcement: detection of illicit material in seized video, …

2012: 56 concepts evaluated

3 Airplane
4 Airplane_Flying
9 Basketball
13 Bicycling
15 Boat_Ship
16 Boy
17 Bridges
25 Chair
31 Computers
51 Female_Person
54 Girl
56 Government_Leader
57 Greeting
63 Highway
71 Instrumental_Musician
72 Kitchen
74 Landscape
75 Male_Person
77 Meeting
80 Motorcycle
84 Nighttime
85 Office
95 Press_Conference
99 Roadway_Junction
101 Scene_Text
105 Singing
107 Sitting_down
112 Stadium
116 Teenagers
120 Throwing
128 Walking_Running
155 Apartments
163 Baby
198 Civilian_Person
199 Clearing
254 Fields
267 Forest
274 George_Bush
276 Glasses
297 Hill
321 Lakes
338 Man_Wearing_A_Suit
342 Military_Airplane
359 Oceans
434 Skier
440 Soldiers

901 Beach + Mountain
902 Old_people + Flags
903 Animal + Snow
904 Bird + Waterscape_waterfront
905 Dog + Indoor
906 Driver + Female_Human_face
907 Person + underwater
908 Table + Telephone
909 Two_People + Vegetation
910 Car + Bicycle

Top 10 InfAP scores by feature (of 51 Full runs)



3 Airplane	4 Airplane_flying	9 Basketball	13 Bicycling	15 Boat_ship	**16 Boy**	**17 Bridges**	**25 Chair**	**31 Computers**	**51 Female_ person**	54 Girl	56 Government_ Leader	**57 Greeting**	**63 Highway**	**71 Instrumental_ Musician**
72 Kitchen	74 Landscape	**75 Male_ person**	77 Meeting	**80 Motorcycle**	**84 Nighttime**	85 Office	95 Press_ Conference	99 Roadway_ Junction	**101 Scene_Text**	105 Singing	**107 Sitting_down**	**112 Stadium**	116 Teenagers	120 Throwing
128 Walking_ Running	155 Apartments	**163 Baby**	198 Civilian_Person	199 Clearing	254 Fields	267 Forest	274 George_Bush	276 Glasses	297 Hill	321 Lakes	338 Man_Wearing_ A_Suit	342 Military_Airplane	359 Oceans	434 Skier
440 Soldiers														

Judges viewed 282,949 clips returned by the systems.

Best systems had on avg. 7 targets in top 10, 70 in top 100

Tech transfer example

Euvision Technologies makes Amsterdam U. semantic indexing software commercially available as Impala *

Thu, 2 Feb 2012

Mr. Over:

We expect to sign our first paid licensing agreement next week. Licensee will be a system integrator who then makes the software available to all police departments in the Netherlands. Concepts to detect are nudity, babies, and children. Application is detection of child abuse in images/videos on confiscated computers/DVDs/tapes.

Your work will have impact on society, in a good way.

Kind regards,

Harro Stokman.

CEO Euvision Technologies,

M: +31 6 41 51 95 67

www.euvt.eu

Matrix II / Science Park 400

1098 XH Amsterdam

NetherlandsEuvision Technologies -/- Premier Visual Concept Detection

* Identification is not intended to imply recommendation or endorsement by NIST

Euvision technologies – customer applications *

- Gerrit Baarda, CEO of Ziuz (www.ziuz.com), says: “We have licensed Impala and integrated it into VizX2. VizX 2 is a total solution for analyzing video and photo material confiscated in investigations into sexual child abuse. Our clients love the new filtering technology. They find the illegal stuff faster, with decreased mental stress for the team.?

- Toon Akkermans, CEO of NCIM (www.ncim.com), says: "We have integrated Impala in our Forensic Dashboard. This Dashboard sits on data of the Dutch Forensic Institute (NFI/Xiraf). In several E-discovery cases, we tried to find documents containing invoices in a big pile of data. Existing text based search found a few: only the ones that were tagged as an invoice. Impala found the rest, hidden in huge set of images. Today, we therefore start with Impala based search."

* Identification is not intended to imply recommendation or endorsement by NIST

Instance search

Example use case: *browsing a video archive, you find a video of a person, place, or thing of interest to you, known or unknown, and want to find more video containing the same target, but not necessarily in the same context.*

System task:

- Given a topic (description of what to look for) with:
  - example segmented images of the target (2-6)
  - a target type (PERSON, PLACE, OBJECT)
- Search collection of 74,958 10-second Flickr clips
- Return a list of up to 1000 clips ranked by likelihood that they contain the topic target

Applications:

- Business intelligence: Logo, product detection
- Person, object, locale linking in archive exploration
- Forensic search in surveillance, seized video, ...?

Topic targets (example images)

Topic: **Examples:**

48 **5**

Mercedes star

60 **6**

Stephen Colbert

68 **6**

Puma logo

56 **9**

Pantheon interior

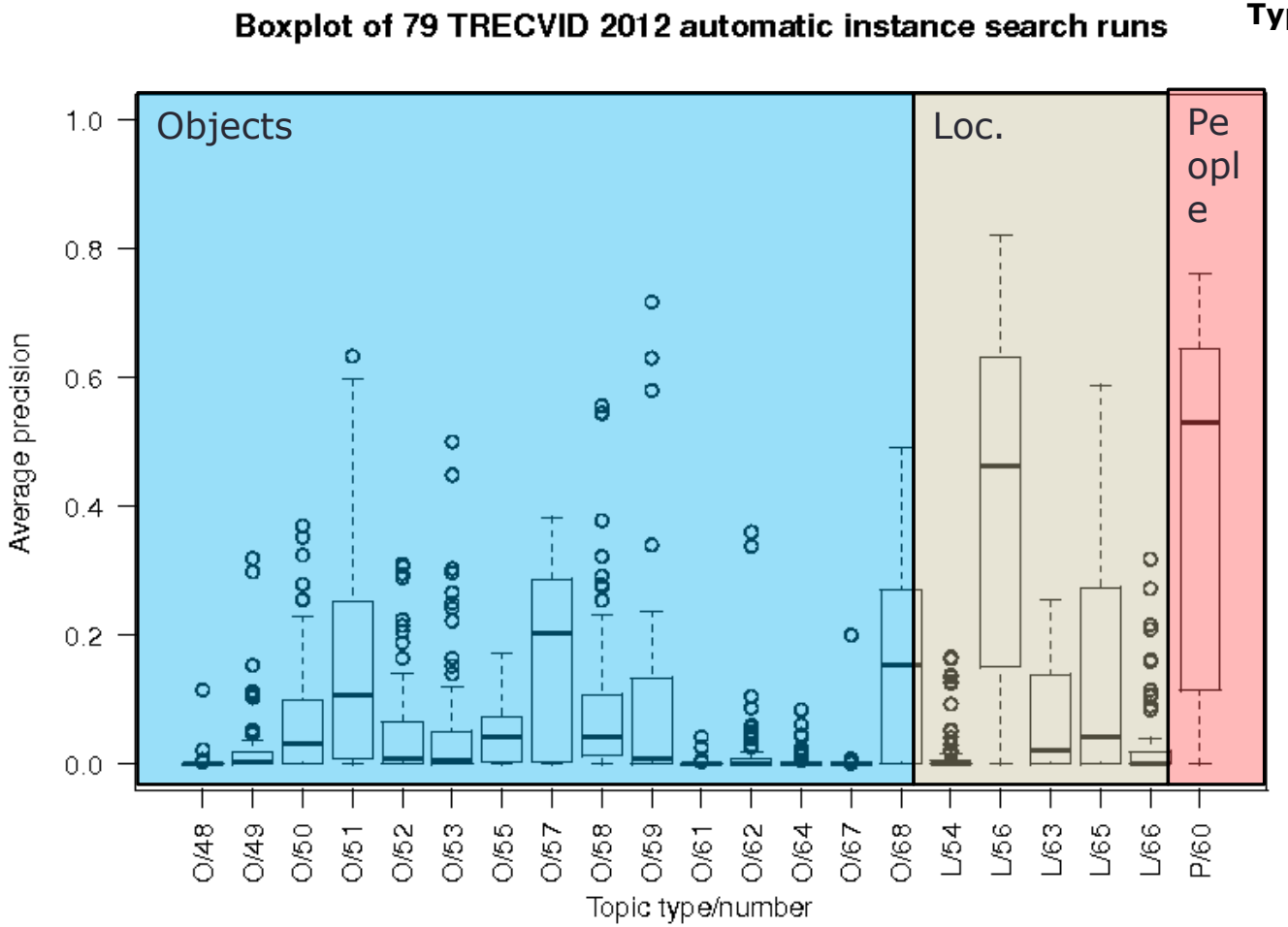
52 **4**

London Underground logo

53 **6**

Coca-Cola logo

Evaluation – results by topic/type - automatic



Type/# Name [clips with target]

O/48 Mercedes star [50]
O/49 Brooklyn bridge tower [57]
O/50 Eiffel Tower [276]
O/51 Golden Gate Bridge [295]
O/52 London Underground logo [59]
O/53 Coca-cola logo (letters) [61]
O/55 Sears/Willis Tower [36]
O/57 Leshan Giant Buddha [21]
O/58 US Capitol exterior [79]
O/59 St. Peter's baldachine [30]
O/61 Pepsi logo – circle [13]
O/62 One World Trade Center [9]
O/64 Empire State Building [51]
O/67 MacDonald's arches [5]
O/68 Puma logo – animal [15]

L/54 Stonehenge [40]
L/56 Pantheon interior [43]
L/63 Prague Castle interior [25]
L/65 Hagia Sophia interior [10]
L/66 Hoover Dam exterior [12]

P/60 Stephen Colbert [45]

NIST judges watched 189,418 clips submitted by systems

Best systems retrieved on average 5 targets in top 10

Content-based copy detection

Example use case: *automatically determine whether a given video contains a (transformed) segment of any reference video (e.g., proprietary, illicit, etc.)*

Test Data:

- Create 201 base clips (2/3 containing reference video)
- Apply each combination of 8 video & 7 audio transformations to the base clips to create 11256 test clips (7*8*201)

System task:

- Given a 400 hour reference collection of Internet videos and 11256 test clips
- Determine for each test clip **whether** it contains reference video and if so, **where** that reference video begins and ends in the test clip

Applications:

- Copyright control (e.g. MovieLabs)
- Business intelligence (advertisement tracking)
- Law enforcement investigations involving specific video

Transformations

Video

- Simulated camcording
- Picture in picture
- Insertions of pattern
- Strong re-encoding
- Change of gamma
- Decrease in quality
- Post production
- Combination of 3 randomly selected

Audio

- Nothing
- mp3 compression
- mp3 compression and multiband companding
- Bandwidth limit and single-band companding
- Mix with speech
- Mix with speech, then multiband compress
- Bandpass filter, mix with speech, compress

Evaluation

- Automatic measurment against ground truth created with the test clips
- **Measures**:
  - weighted sum of miss probability + false alarm rate
  - location accuracy
  - Processing time
- For the test clips that underwent a given transformation,

  a top 2011 system could detect on average 126 of 134 copies with a false alarm rate of almost 0.

More information ➔ TRECVID Website

http://trecvid.nist.gov

- Publications
- Past data
- Evaluation tools
- Past guidelines

paul.over@nist.gov